# A Smirnov-Bickel-Rosenblatt theorem for compactly-supported wavelets

Adam D. Bull
Statistical Laboratory
University of Cambridge
a.bull@statslab.cam.ac.uk

**Abstract**

In nonparametric statistical problems, we wish to find an estimator of an unknown function $f$. We can split its error into bias and variance terms; Smirnov, Bickel and Rosenblatt have shown that, for a histogram or kernel estimate, the supremum norm of the variance term is asymptotically distributed as a Gumbel random variable. In the following, we prove a version of this result for estimators using compactly-supported wavelets, a popular tool in nonparametric statistics. Our result relies on an assumption on the nature of the wavelet, which must be verified by provably-good numerical approximations. We verify our assumption for Daubechies wavelets and symlets, with $N = 6, \ldots, 20$ vanishing moments; larger values of $N$, and other wavelet bases, are easily checked, and we conjecture that our assumption holds also in those cases.

## 1 Introduction

In nonparametric statistical problems, such as density estimation, regression, or white noise, we wish to find an estimate $\hat{f}$ of an unknown function $f$ (Tsybakov, 2009). We can measure the accuracy of an estimator $\hat{f}$ by its distance from $f$, $\|\hat{f} - f\|$, where $\|\cdot\|$ is some norm on functions. We can then decompose the error into variance and bias terms,

$$\|\hat{f} - f\| \leq \|\hat{f} - \mathbb{E}\hat{f}\| + \|\mathbb{E}\hat{f} - f\|,$$

where the bias term $\|\mathbb{E}\hat{f} - f\|$ is deterministic, and the variance term $\|\hat{f} - \mathbb{E}\hat{f}\|$ we hope has an asymptotic distribution independent of $f$.

*Mathematics subject classification 2010.* 62G20 (Primary); 62G07, 62G08, 62G15, 65T60 (Secondary)

*Keywords.* nonparametric statistics, compactly-supported wavelets, asymptotic distribution, confidence sets, supremum norm

In density estimation, for the supremum norm on $[0,1]$,

$$\|f\|_\infty := \sup_{x\in[0,1]} |f(x)|,$$

the limiting distribution of a suitably scaled variance term is given by Smirnov (1950) for histograms, and in the classical paper of Bickel and Rosenblatt (1973) for kernel estimates. In both cases, as the sample size $n$ tends to infinity, the variance term approaches a Gumbel distribution,

$$\mathbb{P}\left(A_n\left(\left\|\frac{\hat{f}_n - \mathbb{E}\hat{f}_n}{\sqrt{f}}\right\|_\infty - B_n\right) \leq x\right) \to e^{-e^{-x}},$$

for known sequences $A_n$, $B_n$. This result has been of key importance for a variety of problems in nonparametric statistics.

Wavelets are an increasingly popular statistical tool, allowing a simple theoretical description of nonparametric problems, and a computationally efficient implementation of their solution. Giné and Nickl (2010) establish an equivalent of these Smirnov-Bickel-Rosenblatt theorems for certain wavelet estimators, using a result of Hüsler, Piterbarg, and Seleznjev (2003) on the convergence of cyclostationary Gaussian processes. Giné and Nickl describe the asymptotic distribution of the supremum, on increasing intervals, of the Gaussian process

$$X(x) := \int K(x,t)\, dB_t,$$

where $K$ is a wavelet projection kernel, and $B$ a Brownian motion; they then link this result to the statistical problem considered above.

Their result holds only for wavelets satisfying certain analytic conditions, which the authors demonstrate are satisfied by Battle-Lemarié wavelets having $N \leq 4$ vanishing moments; Giné, Güntürk, and Madych (2011) extend this to larger values of $N$. Past work has not, however, succeeded in establishing results for the most commonly used wavelets, such as Daubechies wavelets and symlets. These wavelets, unlike those of Battle and Lemarié, are compactly supported, allowing the most efficient implementation of statistical procedures. In the following, we demonstrate that the conditions of Giné and Nickl (2010) hold also in these cases, thereby proving a Smirnov-Bickel-Rosenblatt theorem for the most practically relevant wavelet bases.

We work primarily in the white noise model, but also discuss consequences for the density estimation and regression models. We consider wavelet bases both on $\mathbb{R}$, and also on the interval, using the construction of Cohen et al. (1993). In both cases, we show that the variance term again approaches a Gumbel distribution. We also extend a theorem of Hüsler et al. (2003) (as reported in Hüsler, 1999), establishing a uniform convergene result for cyclostationary processes; this allows us to show that convergence to Gumbel occurs uniformly in large values of the level $x$. These results are

used in Bull (2011) to construct adaptive confidence bands for nonparametric statistical problems, and are also of relevance to many other wavelet procedures.

To prove our results, we must first verify an assumption on the wavelet functions, which in general do not have an analytic form. We therefore make use of provably-accurate numerical approximations, given by Rioul (1992); these approximations also provide an efficient means of computing the constants in our results. We verify our assumption for Daubechies wavelets and symlets, having $N = 6, \dots, 20$ vanishing moments (Daubechies, 1992, §6.4); however, the numerical approximations can easily be applied to larger values of $N$, and other wavelet bases, and we conjecture that our assumption holds also in these cases.

We state our result in Section 2, and describe the necessary numerical approximations in Section 3. We give proofs in Appendix A, and source code in Appendix B.

## 2 Results

To begin, we will need $\varphi$ and $\psi$, the scaling function and wavelet of an orthonormal multiresolution analysis on $L^2(\mathbb{R})$. (For an introduction to wavelets and their statistical applications, see Härdle et al., 1998.) We make the following assumptions on $\varphi$ and $\psi$, which are satisfied, for example, by Daubechies wavelets and symlets, with $N \geq 6$ vanishing moments (Daubechies, 1992, §6.1; Rioul, 1992, §14).

**Assumption 2.1.**

*(i) For $K \in \mathbb{N}$, $\varphi$ and $\psi$ are supported on the interval $[1-K, K]$.*

*(ii) For $N \in \mathbb{N}$, $\psi$ has $N$ vanishing moments:*

$$\int_{\mathbb{R}} x^i \psi(x)\, dx = 0, \qquad i = 0, \dots, N-1.$$

*(iii) $\varphi$ is twice continuously differentiable.*

We will consider wavelet bases on both $\mathbb{R}$ and $[0,1]$, constructed from $\varphi$ and $\psi$. On $\mathbb{R}$, we have an orthonormal basis of $L^2(\mathbb{R})$ given by

$$\varphi_{j_0,k}(x) \coloneqq 2^{j_0/2}\varphi(2^{j_0}x - k), \qquad \psi_{j,k} \coloneqq 2^{j/2}\psi(2^j x - k), \tag{2.1}$$

for some lower resolution level $j_0 \in \mathbb{Z}$, $j > j_0$, and $k \in \mathbb{Z}$.

On $[0,1]$, we can generate an orthonormal basis of $L^2([0,1])$ using the construction of Cohen et al. (1993) (see also Chyzak et al., 2001). We obtain basis functions

$$\varphi_{j_0,k}, \qquad k = 0, \dots, 2^{j_0} - 1,$$

and

$$\psi_{j,k}, \qquad j > j_0,\ k = 0, \dots, 2^j - 1.$$

For $k \in [N, 2^j - N)$, these functions are given by (2.1). For other values of $k$, the basis functions are specially constructed, so as to form an orthonormal basis with desired smoothness properties.

We will also need to make an assumption on the precise form of the scaling function $\varphi$. While this assumption is difficult to verify analytically, we will see in the following section it can be tested using provably good numerical approximations.

**Assumption 2.2.** *The 1-periodic function*

$$\sigma_\varphi^2(t) := \sum_{k \in \mathbb{Z}} \varphi(t-k)^2$$

*attains its maximum* $\overline{\sigma}_\varphi^2$ *at a unique point* $t_0 \in [0,1)$, *and* $(\sigma_\varphi^2)''(t_0) < 0$.

Given these assumptions, suppose we have an unknown function $f$, with empirical wavelet coefficients $\alpha_k, \beta_{j,k}$,

$$f := \sum_k \alpha_k \varphi_{j_0,k} + \sum_{j > j_0} \sum_k \beta_{j,k} \psi_{j,k}.$$

Suppose also that we observe the empirical wavelet coefficients

$$\hat{\alpha}_k := \alpha_k + \epsilon_{j_0,k}, \qquad \hat{\beta}_{j,k} := \beta_{j,k} + \epsilon_{j,k}, \tag{2.2}$$

where the $\epsilon_{j,k}$ are i.i.d. $N(0, \sigma^2)$. This is the case in the white noise model, where we observe the process

$$Y_t = \int_0^t f(s)\, ds + n^{-1/2} B_t,$$

for a Brownian motion $B$. The empirical wavelet coefficients

$$\hat{\alpha}_k = \int \varphi_{j_0,k}(t)\, dY_t, \qquad \hat{\beta}_{j,k} = \int \psi_{j,k}(t)\, dY_t,$$

satisfy (2.2) with $\sigma^2 = n^{-1}$ (Härdle et al., 1998, §10). The model (2.2) also serves as a limiting approximation in density estimation and regression, which we return to later.

The wavelet projection estimate of $f$, at resolution level $j$, is then

$$\hat{f}(j) := \sum_k \hat{\alpha}_k \varphi_{j_0,k} + \sum_{j_0 < l \leq j} \sum_k \hat{\beta}_{l,k} \psi_{l,k}.$$

Set

$$\upsilon_\varphi := -\frac{\sum_{k \in \mathbb{Z}} \varphi'(t_0 - k)^2}{\overline{\sigma}_\varphi \sigma_\varphi''(t_0)}, \tag{2.3}$$

and define the quantities

$$\begin{aligned}
a(j) &\coloneqq \sqrt{2\log(2)j}, \\
b(j) &\coloneqq a(j) - \frac{\log(\pi\log 2) + \log j - \frac{1}{2}\log(1+\upsilon_\varphi)}{2a(j)}, \\
c(j) &\coloneqq \frac{\overline{\sigma}_\varphi}{\sigma} 2^{j/2}, \\
x(\gamma) &\coloneqq -\log\left(-\log(1-\gamma)\right).
\end{aligned}$$

We then have the following result on the distribution of the variance term.

**Theorem 2.3.** *Let $j_n \to \infty$, $\gamma_0 \in (0,1)$, and either:*

*(i) for a wavelet basis on $\mathbb{R}$, $\Gamma_n \coloneqq (0, \gamma_0]$; or*

*(ii) for a wavelet basis on $[0,1]$, $\Gamma_n \coloneqq [\gamma_n, \gamma_0]$, where $\gamma_n \in (0, \gamma_0)$, and $\gamma_n^{-1} = o(e^{Cj_n})$ for any $C > 0$.*

*Then, as $n \to \infty$,*

$$\sup_{\gamma\in\Gamma_n} \left| \gamma^{-1}\mathbb{P}\left( \|\hat{f}(j_n) - \mathbb{E}\hat{f}(j_n)\|_\infty > c(j_n)\left(\frac{x(\gamma)}{a(j_n)} + b(j_n)\right)\right) - 1 \right| \to 0.$$

While this result is stated for the white noise model, similar results hold also in density estimation and regression. In density estimation, $f$ is a density, and we observe

$$X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} f.$$

This can be linked to the white noise model using Giné and Nickl (2010, §4.1). In regression, we have independent observations

$$Y_i \sim N(f(x_i), \sigma^2),$$

for $x_i \coloneqq i/n$, $i = 1, \dots, n$. Regression is known to be asymptotically equivalent to white noise, as in Brown and Low (1996). We can thus transfer our result also to these models.

## 3 Numerical approximations

To apply our result, we must first verify Assumption 2.2, which depends on the function $\varphi$ and its derivatives. In general, $\varphi$ has no explicit form, but we can approximate it numerically using the cascade algorithm. $\varphi$ satisfies a two-scale relation,

$$\varphi(x) = \sum_{k=0}^{2K-1} u_k^{(0)} \varphi(2x + K - k),$$

for filter coefficients $u_0^{(0)}, \dots, u_{2K-1}^{(0)} \in \mathbb{R}$ satisfying

$$\sum_{k \text{ odd}} u_k^{(0)} = \sum_{k \text{ even}} u_k^{(0)} = 1,$$

and we can use these filter coefficients to compute an approximation to $\varphi$.

For $n \geq 0$, $k = 0, \dots, 2K-1$, set

$$u_k^{(n+1)} := 2 \sum_{i=0}^{k} (-1)^i u_{k-i}^{(n)},$$

and for $j, n \geq 0$, $0 \leq k < 2^j(2K-1)$, $x \in [1-K, K)$,

$$g_{0,k}^{(n)} := \delta_k, \qquad g_{j+1,k}^{(n)} := \sum_{i \in \mathbb{Z}} g_{j,i}^{(n)} u_{k-2i}^{(n)},$$

$$f_{j,k}^{(n)} := \sum_{i=0}^{n} \binom{n}{i} (-1)^i g_{j,k-2^j i}^{(n)}, \qquad f_j^{(n)}(x) := f_{j,\lfloor 2^j(x+K-1) \rfloor}^{(n)}.$$

The functions $f_j^{(0)}$ then converge to a limit function $f$ defined by the $u_k^{(0)}$, and the $f_j^{(n)}$ likewise converge to $f^{(n)}$. The following theorem bounds the error in this approximation, and is a straightforward consequence of results in Rioul (1992).

**Theorem 3.1.** *For integers $j, n \geq 0$, set*

$$\alpha_j^{(n)} := 1 - j^{-1} \log_2 \left( \max_{k=0}^{2^j-1} \sum_{i=0}^{2K-2} |g_{j,k+2^j i}^{(n+1)}| \right),$$

$$C_j^{(n)} := \left(1 - 2^{-\alpha_j^{(n)}}\right)^{-1} \left( \max_{l=0}^{j-1} \max_{k=0}^{2^l(2K-1)-1} 2^{(\alpha_j^{(n)}-1)l} |f_{l,k}^{(n+1)}| \right) \left( \max_{m=0,1} \sum_{k=0}^{K-1} \left| \sum_{i=0}^{k} u_{2i+m}^{(n)} - 1 \right| \right).$$

*If $\alpha_j^{(0)} > 0$ for some $j$, the functions $f_j^{(0)}$ converge in $L^\infty$ to a function $f : [1-K, K) \to \mathbb{R}$ satisfying*

$$f(x) = \sum_{k=0}^{2K-1} u_k^{(0)} f(2x + K - k).$$

*If also $\alpha_j^{(n)} > 0$ for some $j$, and $n > 0$, then $f$ is $n$-times-differentiable, and the $f_j^{(n)}$ converge in $L^\infty$ to $f^{(n)}$. For $n \geq 0$, the approximations $f_j^{(n)}$ converge at a rate*

$$\|f_j^{(n)} - f^{(n)}\|_{L^\infty} \leq C_j^{(n)} 2^{-j\alpha_j^{(n)}}.$$

*Furthermore, given integers $j \geq 0$, $a \leq b$, set*

$$I \coloneqq 2^{-j}[a, b+1) + \mathbb{Z}, \qquad J(l) \coloneqq \left[\lfloor 2^{l-j} a \rfloor - 2K + 2, \lfloor 2^{l-j} b \rfloor\right] + 2^l \mathbb{Z}.$$

*Then, on $I \cap [1-K, K)$:*

*(i) for $l \geq j$, the values of $f_l^{(n)}$ depend on $g_{j,k}^{(n)}$ only for $k \in J(j)$; and*

*(ii) the above results hold also for quantities $\alpha_j^{(n)}(I)$ and $C_j^{(n)}(I)$ defined similarly, taking maxima over $g_{l,k}^{(n+1)}$ and $f_{l,k}^{(n+1)}$ only for $k \in J(l)$.*

The function $\varphi$ may be defined as the limit of this procedure (Daubechies, 1992, §6.5). We may thus compute upper and lower bounds on $\varphi$ and its derivatives. Note that, while we could obtain values of the derivatives by finite differencing, this would be numerically unstable, and lead to poor bounds; the above procedure provides good bounds on all derivatives of $\varphi$.

To verify Assumption 2.2, and to compute the constants $\overline{\sigma}_\varphi^2$ and $\upsilon_\varphi$, we must use these bounds to control the function $\sigma_\varphi^2$, and its derivatives. Doing so over the whole of $[0, 1]$ requires memory exponential in $j$, which quickly becomes infeasible. However, once we have approximated $\sigma_\varphi^2$ well enough to know that its maxima lie in some interval $I$, we can exploit the local nature of the cascade algorithm, and its bounds, to approximate $\sigma_\varphi^2$ only over $I$. As the resolution $j$ increases, so does the accuracy with which we can locate the maxima, ensuring our memory costs remain manageable.

In our implementation, we choose $I$ to be the smallest interval containing all points $t$ for which the bounds on $\sigma_\varphi^2$, and its derivative, are consistent with:

(i) $\sigma_\varphi^2(t) = \sup_{s \in [0,1]} \sigma_\varphi^2(s)$; and

(ii) $(\sigma_\varphi^2)'(t) = 0$.

Note that to ensure efficiency, we must allow choices of $I$ which wrap around the edges of $[0, 1]$; in other words, we must allow $I$ to be any interval on the torus. If we find an interval $I$ containing all maxima of $\sigma_\varphi^2$, with the property that $\sigma_\varphi'' \leq -\varepsilon < 0$ on $I$, we may conclude Assumption 2.2 is satisfied. We have thus described Algorithm 1.

To obtain high accuracy, the computation of the filter coefficients $u_k$, and subsequent approximations, must be performed using variable-precision arithmetic; the rounding error in these computations must likewise be controlled with interval arithmetic. We satisfy these requirements by implementing the above algorithm in the computer algebra system Mathematica. For Daubechies wavelets and symlets, $N = 6, \ldots, 20$, we find that Assumption 2.2 is indeed satisfied, and obtain accurate values of $\overline{\sigma}_\varphi^2$ and $\upsilon_\varphi$, given in Table 1.

**Algorithm 1** Verify assumption and compute constants

$I \leftarrow [0, 1]$
$j \leftarrow 0$
**repeat**
  calculate approximations $\varphi_j^{(n)}$ to $\varphi^{(n)}$ on $I$, $n = 0, 1, 2$
  deduce bounds on $(\sigma_\varphi^2)^{(n)}$ on $I$, $n = 0, 1, 2$
  deduce bounds on $\overline{\sigma}_\varphi^2$ and $\upsilon_\varphi$
  $I \leftarrow$ smallest interval known to contain all maxima of $\sigma_\varphi^2$
  $j \leftarrow j + 1$
**until** desired accuracy reached
**if** $\sigma_\varphi''$ bounded below zero on $I$ **then**
  Assumption 2.2 is verified
**end if**

| | Daubechies | | Symlet | |
|---|---|---|---|---|
| $N$ | $\overline{\sigma}_\varphi^2$ | $\upsilon_\varphi$ | $\overline{\sigma}_\varphi^2$ | $\upsilon_\varphi$ |
| 6 | 1.251 716 | 0.221 993 | 1.361 961 | 0.106 518 |
| 7 | 1.276 330 | 0.197 328 | 1.253 835 | 0.248 681 |
| 8 | 1.250 928 | 0.266 316 | 1.286 722 | 0.173 642 |
| 9 | 1.222 637 | 0.275 519 | 1.232 334 | 0.302 351 |
| 10 | 1.199 772 | 0.391 629 | 1.243 114 | 0.255 337 |
| 11 | 1.195 384 | 0.415 019 | 1.209 007 | 0.324 200 |
| 12 | 1.189 984 | 0.445 388 | 1.215 480 | 0.335 022 |
| 13 | 1.182 351 | 0.460 792 | 1.195 567 | 0.385 147 |
| 14 | 1.172 690 | 0.510 179 | 1.195 969 | 0.405 884 |
| 15 | 1.165 335 | 0.553 767 | 1.184 307 | 0.446 419 |
| 16 | 1.159 678 | 0.594 027 | 1.181 901 | 0.465 670 |
| 17 | 1.154 955 | 0.621 941 | 1.174 105 | 0.496 485 |
| 18 | 1.150 103 | 0.652 913 | 1.170 871 | 0.520 228 |
| 19 | 1.145 393 | 0.686 434 | 1.164 974 | 0.551 765 |
| 20 | 1.141 050 | 0.722 113 | 1.161 837 | 0.571 150 |

Table 1: Computed values of constants

## Acknowledgements

We would like to thank Richard Nickl for his valuable comments and suggestions.


# A Proofs

We will need the following result, which is a version of Theorem 1 in Hüsler (1999). The result concerns the maxima of centred Gaussian processes whose variance functions are periodic; such processes are called *cyclostationary*. In Hüsler's original result, the maxima of a sequence of processes was shown to converge to a Gumbel random variable. In our result, we will specialise to a single process, and show this convergence occurs uniformly.

**Lemma A.1.** *Let $T = T(n) \to \infty$ as $n \to \infty$. In the notation of Hüsler (1999), let (A1)–(A3) and (B1)–(B4) hold, for a fixed process $X_n(t) = X(t)$, not depending on $n$. Further let $\alpha = \beta$, and let Hüsler's condition (1) hold. Define*

$$u(\tau) = \sigma_n \mu^{-1}(\tau / m_T).$$

*Then for any $\tau_0 > 0$, we have*

$$\sup_{\tau \in (0,\tau_0]} \left| \frac{\mathbb{P}(M_n(T) > u(\tau))}{1 - e^{-\tau}} - 1 \right| \to 0$$

*as $n \to \infty$.*

*Proof.* Our argument proceeds as in the proof of Theorem 1 in Hüsler (1999). Without loss of generality, we may assume that $\sigma_n = 1$. For $\tau \leq \tau_0$, $u(\tau) \geq u(\tau_0) \to \infty$, and by definition

$$m_T \mu(u(\tau)) = \tau.$$

The approximation errors in parts (i) and (ii) of Hüsler's proof are thus $O(g(S)\tau)$ and $O(\rho_c \tau)$ respectively. In part (iii), we note that

$$u(\tau)^2 = 2 \log(T/\tau) - \log \log(T/\tau) + O(1),$$

so Hüsler's term (4) is of order

$$\begin{aligned} &\tau^{1+\eta}(T/\tau)^{1+\eta} u(\tau)^{2/\alpha} \exp\{-u(\tau)^2/(1+\gamma)\} \\ &\qquad = \tau^{1+\eta} \exp\{-[(1-\gamma)/(1+\gamma) - \eta] \log(T/\tau) + o(\log(T/\tau))\} = o(\tau), \end{aligned}$$

and term (5) is of order

$$\begin{aligned} &\tau^2 (T/\tau)^2 \delta(T^\eta) \exp\{-(2\log(T/\tau) - \log\log(T/\tau))(1 - \delta(T^\eta)\} \\ &\qquad\qquad = O(\tau^2 \delta(T^\eta) \log(T/\tau)) = o(\tau). \end{aligned}$$

In Hüsler's final display, we may thus write

$$\mathbb{P}(M_n(T) \le u(\tau)) = \exp\{-(1+o(1))\tau\} + o(\tau).$$

As the process $X(t)$ does not depend on $n$, the error in each of these approximations depends only on $u = u(\tau)$, and the above limits hold as $u \to \infty$. (This can be seen from the precise form of the errors, as given in Piterbarg and Seleznjev, 1994, §3.1, and in Hüsler's proof.) Since $u$ is decreasing in $\tau$, the limits are therefore uniform in $\tau$ small.

Consider the function

$$f(x, y; \tau) \coloneqq \log\left(\frac{1 - \exp(-(1+x)\tau)}{\tau} + y\right),$$

defined on $0 \le \tau \le \tau_0$, $|x| \le \frac{1}{2}$, $|y| \le \frac{1}{2}(1 - \exp(-\frac{1}{2}\tau_0))/\tau_0$. The derivatives

$$\frac{\delta f}{\delta x} = \frac{\exp(-(1+x)\tau)}{\exp f}, \qquad \frac{\delta f}{\delta y} = \frac{1}{\exp f}$$

are finite, and continuous in $x$, $y$ and $\tau$, so by the mean value inequality, for $n$ large,

$$\begin{aligned}\log\left(\frac{\mathbb{P}(M_n(T) > u(\tau))}{\tau}\right) &= f(o(1), o(1); \tau)\\ &= f(0, 0; \tau) + o(1)\\ &= \log\left(\frac{1 - e^{-\tau}}{\tau}\right) + o(1).\end{aligned}$$

As the above limits are uniform in $\tau \le \tau_0$, the result follows. □

We now apply this result to a cyclostationary process, composed of scaling functions $\varphi$, which we can use to model the variance of estimators $\hat{f}(j_n)$.

**Lemma A.2.** *Define the cyclostationary Gaussian process*

$$X(t) \coloneqq \overline{\sigma}_\varphi^{-1} \sum_{k \in \mathbb{Z}} \varphi(t-k) Z_k, \qquad Z_k \overset{\text{i.i.d.}}{\sim} N(0,1).$$

*For any* $\gamma_0 \in (0,1)$, $j_n \to \infty$,

$$\sup_{\gamma \in (0, \gamma_0]} \left| \gamma^{-1} \mathbb{P}\left( \sup_{t \in [0, 2^{j_n}]} |X(t)| > \frac{x(\gamma)}{a(j_n)} + b(j_n) \right) - 1 \right| \to 0$$

*as* $n \to \infty$.

*Proof.* For fixed $\gamma$, the result is a consequence of Theorem 2 in Giné and Nickl (2010); the statement uniform over $(0, \gamma_0]$ follows, replacing Theorem 1 of Hüsler (1999) in Giné and Nickl's proof with Lemma A.1. The conditions of Giné and Nickl's theorem are satisfied by Assumptions 2.1 and 2.2, as follows.

(i) $X$ has almost-sure derivative

$$X'(t) := \overline{\sigma}_\varphi^{-1} \sum_{k\in\mathbb{Z}} \varphi'(t-k)Z_k,$$

so is continuous. $X'$ is also the mean square derivative:

$$\begin{aligned} &h^{-1}\mathbb{E}[(X(t+h)-X(t)-hX'(t))^2] \\ &\qquad\qquad = h^{-1}\sum_{k\in\mathbb{Z}} \left(\varphi(t-h-k)-\varphi(t-k)-h\varphi'(t-k)\right)^2, \end{aligned}$$

which tends to 0 as $h \to 0$, since the sum has finitely many non-zero terms.

(ii) For $i = 0, 1$, define functions $f_i(x) := x^i$ on $[0, 1]$, having wavelet expansions

$$f_i = \sum_k \alpha^i_{J,k}\varphi_k + \sum_{j>J}\sum_k \beta^i_{j,k}\psi_{j,k}$$

in our wavelet basis on $[0, 1]$, for some $J \geq j_0$, $2^J \geq 6K$. As $\psi$ is twice continuously differentiable, and $\varphi$ and $\psi$ have compact support, by Corollary 5.5.4 in Daubechies (1992), $\psi$ has at least two vanishing moments. Thus

$$\beta^i_{j,k} = \langle x^i, \psi_{j,k}\rangle = 0,$$

and

$$f_i(t) = \sum_k \alpha^i_k \varphi_{J,k}(t).$$

For $t \in [0, 1]$, let $v(t)$ denote the vector $(\varphi_{J,k}(t)) \in \mathbb{R}^{2^J}$, so $f_i(t) = \langle \alpha^i, v(t)\rangle$. Given $s \neq t$, we have

$$\begin{aligned} \langle \alpha^0, v(s)\rangle &= 1 = \langle \alpha^0, v(t)\rangle, \\ \langle \alpha^1, v(s)\rangle &= s \neq t = \langle \alpha^1, v(t)\rangle, \end{aligned}$$

so the vectors $v(s)$, $v(t)$ are linearly independent.

For $s, t \in \mathbb{R}$, define

$$r_X(s,t) := \mathbb{C}\mathrm{ov}[X(s), X(t)], \qquad \sigma^2_X(t) := \mathbb{V}\mathrm{ar}[X(t)] = r_X(t,t).$$

Then, if $s, t \in [-K, K]$,

$$\begin{aligned} r_X(s,t) &= \overline{\sigma}_\varphi^{-2} \sum_{k\in\mathbb{Z}} \psi(s-k)\psi(t-k) \\ &= \overline{\sigma}_\varphi^{-2} 2^{-J} \langle v(\tfrac{1}{2} + 2^{-J}s), v(\tfrac{1}{2} + 2^{-J}t)\rangle, \end{aligned}$$

so by Cauchy-Schwarz,

$$r_X(s,t)^2 < \sigma_X^2(s)\sigma_X^2(t).$$

If $s, t \in [k-K, k+K]$ for some $k \in \mathbb{Z}$, the same applies by cyclostationarity. If not, then as $\varphi$ is supported on $[1-K, K]$, we have $r_X(s,t) = 0$. However, for any $t \in [0,1]$, $\langle \alpha^1, v(\frac{1}{2} + 2^{-J}t)\rangle = 1$, so

$$\sigma_X^2(t) = \overline{\sigma}_\varphi^{-2} 2^{-J} \|v(t)\|^2 > 0,$$

and by cyclostationarity the same holds for all $t \in \mathbb{R}$. We thus again obtain

$$r_X(s,t)^2 < \sigma_X^2(s)\sigma_X^2(t).$$

(iii) We have

$$\sigma_X^2(t) = \overline{\sigma}_\varphi^{-2}\sigma_\varphi^2(t),$$

so by Assumption 2.2, $\sup_{t\in[0,1]} \sigma_X^2(t) = 1$, and this maximum is attained at a unique $t_0 \in [0,1)$. If $t_0 \in (0,1)$, this satisfies the conditions of the theorem directly; if not we may proceed as in Proposition 9 of Giné and Nickl (2010). $\sigma_\varphi^2$ is twice differentiable,

$$2\overline{\sigma}_\varphi \sigma_X'(t_0) = (\sigma_\varphi^2)'(t_0)\sigma_\varphi^2(t_0)^{-1/2} = 0,$$

and

$$\begin{aligned} 2\overline{\sigma}_\varphi \sigma_X''(t_0) &= (\sigma_\varphi^2)''(t_0)\sigma_\varphi^2(t_0)^{-1/2} - \tfrac{1}{2}(\sigma_\varphi^2)'(t_0)^2\sigma_\varphi^2(t_0)^{-3/2} \\ &= (\sigma_\varphi^2)''(t_0)\sigma_\varphi^2(t_0)^{-1/2} < 0. \end{aligned}$$

Finally, let $v'(t)$ denote the vector $(\varphi_{J,k}'(t)) \in \mathbb{R}^{\mathbb{Z}}$. Then for $t \in [0,1]$,

$$\langle \alpha^1, v'(t)\rangle = f_1'(t) = 1,$$

so

$$\begin{aligned} \mathbb{E}[X'(t_0)^2] &= \overline{\sigma}_\varphi^{-2} \sum_{k\in\mathbb{Z}} \varphi'(t_0-k)^2 \\ &= \overline{\sigma}_\varphi^{-2} 2^{-J} \|v'(\tfrac{1}{2} + 2^{-J}t_0)\|^2 > 0. \end{aligned} \tag{A.1}$$

(iv) Since $\varphi$ has support $[1-K, K]$,

$$\sup_{s,t:|s-t|\geq 2K-1} |r_X(s,t)| = 0. \qquad \square$$

We may now bound the variance of $\hat{f}(j_n)$. We will show that the variance process is distributed as the process $X$ from the above lemma, so can be controlled similarly.

*Proof of Theorem 2.3.* Let $I_n \coloneqq [0, 2^{j_n}]$. The process

$$X_n(t) \coloneqq \frac{\hat{f}(j_n) - \bar{f}(j_n)}{c(j_n)}(2^{-j_n}t), \qquad t \in I_n,$$

is distributed as

$$\overline{\sigma}_\varphi^{-1} 2^{-j_n/2} \left( \sum_{k\in\mathbb{Z}} Z_{j_0,k}\varphi_{j_0,k}(2^{-j_n}t) + \sum_{j=j_0+1}^{j_n} \sum_{k\in\mathbb{Z}} Z_{j,k}\psi_{j,k}(2^{-j_n}t) \right),$$

for $Z_{j,k} \overset{\text{i.i.d.}}{\sim} N(0,1)$, so by an orthogonal change of basis, as

$$\overline{\sigma}_\varphi^{-1} 2^{-j_n/2} \sum_{k\in\mathbb{Z}} Z_k \varphi_{j_n,k}(2^{-j_n}t), \qquad Z_k \overset{\text{i.i.d.}}{\sim} N(0,1).$$

In case (i), $X_n$ is distributed as the process $X$ from Lemma A.2, so we are done.

In case (ii), set $J_n \coloneqq [2K, 2^{j_n} - 2K]$, and $K_n \coloneqq I_n \setminus J_n$. On $J_n$, $X_n$ is distributed as the process $X$ from Lemma A.2, and we have

$$\begin{aligned}
\mathbb{P}\left(\sup_{t\in J_n} |X_n(t)| > u\right) &\le \mathbb{P}\left(\sup_{t\in I_n} |X_n(t)| > u\right) \\
&\le \mathbb{P}\left(\sup_{t\in J_n} |X_n(t)| > u\right) + \mathbb{P}\left(\sup_{t\in K_n} |X_n(t)| > u\right),
\end{aligned}$$

so for $u_n(j_n) \coloneqq x(\gamma_n)/a(j_n) + b(j_n)$,

$$\begin{aligned}
&\left| \mathbb{P}\left(\sup_{t\in I_n} |X_n(t)| > u_n(j_n)\right) - \mathbb{P}\left(\sup_{t\in J_n} |X(t)| > u_n(j_n)\right) \right| \\
&\qquad\le \mathbb{P}\left(\sup_{t\in K_n} |X_n(t)| > u_n(j_n)\right) \\
&\qquad\le 8K(1 - \Phi(C u_n(j_n))) \\
&\qquad\lesssim e^{-C^2 u_n(j_n)^2/2}/u_n(j_n),
\end{aligned}$$

with a constant $C > 0$ depending on $\varphi$. This term is $o(\gamma_n)$, so the result follows by Lemma A.2, applied to the process $X$ on $J_n$. $\square$

# B Source code

The following program implements Algorithm 1 in Mathematica 8 or above. Note that we bound $\upsilon_\varphi$ by bounding the numerator and denominator of (2.3) separately over $I$. By (A.1), the numerator is positive; to bound $\upsilon_\varphi$ inside $(0, \infty)$, we must therefore bound $\sigma''_\varphi$ below zero. To verify Assumption 2.2, it is thus sufficient that we establish a finite positive value of $\upsilon_\varphi$.

```
Main::usage =
  "Main[w, p, d, jmax, kmax] = bounds on the parameters sigma^2_phi and\n" <>
  "upsilon_phi for the Wavelet w, performing computations using p digits\n" <>
  "of accuracy, stopping once accurate to d digits, after jmax steps,\n" <>
  "after evaluating phi at 2^kmax locations simultaneously, or after\n" <>
  "the results become indeterminate due to lack of precision.\n\n" <>
  "Main[DaubechiesWavelet[6], 200, 6, 100, 12] = \n" <>
  "  {Interval[{1.251716, 1.251716}], Interval[{0.221993, 0.221993}]}"
Main[wavelet_, precision_, digits_, jmax_, kmax_] :=
  Cascade[Filter[wavelet, precision], 2, PhiParams[digits, jmax, kmax]]

Filter::usage =
  "Filter[w] = the cascade filter of a Wavelet w."
Filter[w_, p_:MachinePrecision] :=
  2 Transpose[WaveletFilterCoefficients[w, WorkingPrecision -> p]][[2]]

PhiParams::usage =
  "PhiParams[d, jm, km] = a callback for use with Cascade, which\n" <>
  "computes the parameters sigma^2_phi and upsilon_phi to d d.p., with\n" <>
  "resource limits jm, km."
PhiParams[d_, jm_, km_] :=
  Params[Function[{w, j, offset, phi},
  Module[{PhiMN, s2, s2max, ssp, sp2, sspp, s22b, s22a, ba, zoom},
  PhiMN[m_, n_] := Plus @@ If[m == n, phi[[n+1]]^2, phi[[m+1]] phi[[n+1]]];

  s2 = PhiMN[0, 0]; ssp = PhiMN[0, 1];
  s2max = IntervalMax[s2];
  sp2 = PhiMN[1, 1]; sspp = PhiMN[0, 2];
  s22b = IntervalRange[sp2];
  s22a = -IntervalRange[sspp+sp2];
  ba = s22b / s22a;

  zoom = Thread[(!TrueQ[# < s2max] & /@ s2) && (!TrueQ[# != 0] & /@ ssp)];
  {{s2max, ba}, zoom}]], d, jm, km]

Params::usage =
  "Params[Callback, digits, jmax, kmax] = a callback function for\n" <>
  "use with Cascade. The function will pass its arguments to Callback,\n" <>
  "which should return estimates of parameters, and an array of indices k\n"<>
  "to restrict to. The function will terminate the cascade algorithm if\n" <>
  "the returned parameters are accurate to digits d.p., the resource\n"<>
  "limits jmax, kmax are reached, or the computation is indeterminate.\n" <>
  "Otherwise, it will restrict the algorithm to an interval I containing\n" <>
  "the requested indices, and continue."
Params[Callback_, digits_, jmax_, kmax_] :=
  Function[{w, j, offset, phi}, Module[{ret, zoom, len, i, l},
  {ret, zoom} = Callback[w, j, offset, phi];
  len = Dimensions[phi][[3]];
  Which[
  MemberQ[ret, Indeterminate, Infinity],
    Print["indeterminate, j = ", j];
    ret = If[MemberQ[#, Indeterminate, Infinity],
    Interval[{-Infinity, Infinity}], #] & /@ ret;
    Prepend[ret, True],
  (And @@ (IntervalAccurate[#, digits] &) /@ ret) ||
  (j == jmax && (Print["hit jmax"]; True)) ||
  (len >= 2^kmax && (Print["hit kmax, j = ", j]; True)),
    Prepend[ret, True],
  len == 2^j,
    {l, i} = LongestSubsequence[Join[zoom, zoom], # == False &];
    If[l < 2^(j-1), {False}, Prepend[Mod[#-1, 2^j]+1 & /@ {i+l, i-1}, False]],
  True,
```

```
    Prepend[Through[{Min, Max}[Position[zoom, True]]], False]]]]

InternalAccurate::usage =
  "IntervalAccurate[i, d] = True if the Interval i is accurate to d d.p."
IntervalAccurate[int_, digits_] :=
  Equal @@ Round[Through[{Min, Max}[int]], 10^(-digits)]

IntervalRange::usage =
  "IntervalRange[is] = an Interval giving the range of the Intervals is."
IntervalRange[ints_] :=
  Interval @ Through[{Min, Max}[ints]]

IntervalMax::usage =
  "IntervalMax[is] = an Interval giving the maximum of the Intervals is."
IntervalMax[ints_] :=
  Interval @ Map[Max, Transpose[ints /. Interval -> Identity]]

LongestSubsequence::usage =
  "LongestSubsequence[x, p] = {l, i}, where l is the length of the\n" <>
  "longest consecutive subsequence of x whose members satisfy p, and i\n" <>
  "is the index of its first member."
LongestSubsequence[x_, crit_] :=
  Block[{$RecursionLimit = Infinity}, Module[{m, i, l},
  m = LengthWhile[x, crit];
  If[m == Length[x], {m, 1},
    {l, i} = LongestSubsequence[Drop[x, m + 1], crit];
    If[l > m, {l, i + m + 1}, {m, 1}]]]]

Cascade::usage =
  "Cascade[w, n, Callback] runs the cascade algorithm with filter w,\n" <>
  "bounding f and its first n derivatives.\n\n" <>
  "The function Callback[w, j, offset, phi] should return a list ret.\n" <>
  "If First[ret] is True, cascade will halt, returning Rest[ret].\n" <>
  "If First[ret] is False, the algorithm will continue, and if\n" <>
  "Rest[ret] = {a, b} is given, future evaluation of f will be\n" <>
  "restricted to the indices a,...,b of phi.\n"
Cascade[w_, n_, Callback_] :=
  Module[{offset, u, j, k, g, f, fs, eps, phi, a, b, pts},
  offset = 0; j = 0; k = Length[w]/2;
  u = BuildU[w, n+1]; g = InitG[k, n+1]; fs = {BuildF[g]};
  While[True,
    g = StepG[g, u]; f = BuildF[g]; eps = BoundError[fs, g, u];
    phi = ApplyError[f, eps, k]; AppendTo[fs, f]; offset *= 2; j += 1;
    ret = Callback[w, j, offset, phi]; If[First[ret], Return[Rest[ret]]];
    {a, b} = If[Length[ret] > 1, Rest[ret], {1, Dimensions[phi][[3]]}];
    If[a > b && Dimensions[phi][[3]] == 2^j,
      {fs, g} = WrapFG[fs, g, k]; b = b + 2^j; offset -= 2^j];
    g = Take[g, All, All, {a, 2(k-1)+b}];
    fs = RestrictF[fs, offset, a, b, k]; offset += a-1]]

BuildU::usage =
  "BuildU[u^(0), n] = u^(0:n)"
BuildU[w_, n_] :=
  Map[Function[un, Table[un[[m;;All;;2]], {m, 2}]],
  With[{signs = Table[(-1)^i, {i, Length[w]}]},
  NestList[2 signs Accumulate[signs #] &, w, n]]]

InitG::usage =
  "InitG[k, n] = g_0^(0:n)"
InitG[k_, n_] :=
  ConstantArray[Reverse @ IdentityMatrix[2k-1], n+1]
```

```
StepG::usage =
  "StepG[g_j^(0:n), u^(0:n)] = g_{j+1}^(0:n)"
StepG[g_, u_] :=
  MapThread[Function[{gn, un}, Function[gni,
  Riffle @@ (ListConvolve[#, gni] &) /@ un] /@ gn], {g, u}]

BuildF::usage =
  "BuildF[g_j^(0:n)] = f_j^(0:n)"
BuildF[g_] :=
  MapIndexed[Function[{gn, part}, Module[{n, bn},
  n = First[part]-1; bn = Table[Binomial[n, i](-1)^i, {i, 0, n}];
  Transpose[ListConvolve[bn, #, 1, 0] & /@ Transpose @ gn]]], g]

BoundAlpha::usage =
  "BoundAlpha[g_j^(0:n), j] = alpha_j^(0:n-1)"
BoundAlpha[g_, j_] :=
  1-Log[2, Max @@ Plus @@ Abs @ #]/j & /@ Rest @ g

BoundC::usage =
  "BoundC[f_{0:j-1}^(0:n), u^(0:n), alpha_j^(0:n-1)] = C_j^(0:n-1)"
BoundC[f_, u_, alpha_] :=
  MapThread[Function[{fn, un, an}, If[an <= 0, Infinity,
  (Max @ MapIndexed[2^((an-1)(First[#2]-1)) Abs[#1] &, fn])
  (Max @@ Plus @@ (Abs[Accumulate[#]-1] &) /@ un)/(1-2^(-an))]],
  {Rest @ Transpose @ f, Most @ u, alpha}]

BoundError::usage =
  "BoundError[f_{0:j-1}^(0:n), g_j^(0:n), u^(0:n)] = eps_j^(0:n-1)"
BoundError[f_, g_, u_] := Module[{j, alpha, c},
  j = Length[f]; alpha = BoundAlpha[g, j]; c = BoundC[f, u, alpha];
  MapThread[Function[{cn, an}, cn 2^(-j an)], {c, alpha}]]

ApplyError::usage =
  "ApplyError[f_j^(0:n), eps_j^(0:n-1), k] = intervals bounding f_j^(0:n-1)"
ApplyError[f_, eps_, k_] :=
  MapThread[Function[{fn, en},
  Map[Interval[{#-en, #+en}] &, fn, {2}]],
  {Most @ Take[f, All, All, {2k-1, -1}], eps}]

WrapFG::usage =
  "WrapFG[f_{0:j}^(0:n), g_j^(0:n), k] = f and g wrapped around at integers"
WrapFG[f_, g_, k_] :=
  Module[{Wrap}, Wrap[x_] :=
    Map[Function[xn, Map[Function[xni, Join[xni[[1]], xni[[2]][[2k-1;;]]]],
    Partition[xn, 2, 1, {-1, 1}, {ConstantArray[0, Dimensions[x][[3]]]}]]],
    x]; {Wrap /@ f, Wrap @ g}]

RestrictF::usage =
  "RestrictF[f_{0:j}^(0:n), offset, a, b, k] = f_{0:j}^(0:n)|offset+[a, b]"
RestrictF[fs_, offset_, a_, b_, k_] :=
  MapIndexed[Function[{fj, part},
  Module[{scale, start, end}, scale = 2^(Length[fs]-First[part]);
  start = Floor[(offset+a-1)/scale]-Floor[offset/scale]+1;
  end = 2(k-1)+Floor[(offset+b-1)/scale]-Floor[offset/scale]+1;
  Take[fj, All, All, {start, end}]]], fs]
```

# References


Bickel P J and Rosenblatt M. On some global measures of the deviations of density function estimates. *The Annals of Statistics*, 1:1071–1095, 1973.

Brown L D and Low M G. Asymptotic equivalence of nonparametric regression and white noise. *The Annals of Statistics*, 24(6):2384–2398, 1996. doi:10.1214/aos/1032181159

Bull A D. Honest adaptive confidence bands and self-similar functions. October 2011. arXiv:1110.4985

Chyzak F, Paule P, Scherzer O, Schoisswohl A, and Zimmermann B. The construction of orthonormal wavelets using symbolic methods and a matrix analytical approach for wavelets on the interval. *Experimental Mathematics*, 10(1):67–86, 2001.

Cohen A, Daubechies I, and Vial P. Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis*, 1(1):54–81, 1993. doi:10.1006/acha.1993.1005

Daubechies I. *Ten lectures on wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

Giné E, Güntürk C S, and Madych W R. On the periodized square of $L^2$ cardinal splines. *Experimental Mathematics*, 20(2):177–188, 2011.

Giné E and Nickl R. Confidence bands in density estimation. *The Annals of Statistics*, 38(2):1122–1170, 2010. doi:10.1214/09-AOS738

Härdle W, Kerkyacharian G, Picard D, and Tsybakov A. *Wavelets, approximation, and statistical applications*, volume 129 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1998.

Hüsler J. Extremes of Gaussian processes, on results of Piterbarg and Seleznjev. *Statistics & Probability Letters*, 44(3):251–258, 1999. doi:10.1016/S0167-7152(99)00016-4

Hüsler J, Piterbarg V, and Seleznjev O. On convergence of the uniform norms for Gaussian processes and linear approximation problems. *The Annals of Applied Probability*, 13 (4):1615–1653, 2003. doi:10.1214/aoap/1069786514

Piterbarg V and Seleznjev O. Linear interpolation of random processes and extremes of a sequence of Gaussian nonstationary processes. 1994. Technical report 446, Department of Statistics, University of North Carolina, Chapel Hill, NC.

Rioul O. Simple regularity criteria for subdivision schemes. *SIAM Journal on Mathematical Analysis*, 23(6):1544–1576, 1992. doi:10.1137/0523086

Smirnov N V. On the construction of confidence regions for the density of distribution of random variables. *Doklady Akad. Nauk SSSR (N.S.)*, 74:189–191, 1950.

Tsybakov A B. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer, New York, 2009.